REMARKS

This amendment responds to the Office Action mailed April 7, 2006. In the Office Action the Examiner:

- rejected claims 14, 15, 17, 52, 53, and 55 as being indefinite under 35 U.S.C. 112, second paragraph, as being indefinite;

- rejected claims 12-17, 40-48 and 50-55 under 35 U.S.C. 102(e) as anticipated by Meyerzon et al. (US 6,547,829);

- rejected claims 18-20, 37-39 and 56-58 under 35 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (US 6,547,829) in view of Rujan et al. (US 6,976,207); and

- rejected claim 49 under 37 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (US 6,547,829) in view of Lambert et al. (US Pub. No. 2002/0038350).

After entry of this amendment, the pending claims are: claims 12-20, 37-40 and 42-58.

Applicants have revised claims 14, 15, 17, 52, 53 and 55 to address the rejections under 35 U.S.C. 112. Claim 40 has been revised to incorporate claim 41.

Claim Rejections - 35 U.S.C. § 112

Claims 14, 15, 17, 52, 53, and 55 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. In particular, the Examiner states that insufficient antecedent basis exists in claims 14, 15, 52, 53, and 55 for "the particular document" and for "the requesting document".

To address this rejection, dependent claims 14, and 15 have been amended to change their dependencies from claim 12 to claim 13. Similarly dependent claims 52 and 53 have been amended to change their dependencies from 50 to 51. These changes provide sufficient antecedent basis for "the particular document".

Furthermore, "the requesting document" has been replaced by "the newly crawled document" in claims 14, 15, 17, 52, 53 and 55. Accordingly, it is respectfully submitted that the Examiner's 35 U.S.C. 112 rejections have been addressed. The Applicants respectfully requests that the respective claim rejections be withdrawn.

The Examiner has rejected claims 12-17, 40-48 and 50-55 under 35 U.S.C. 102(e) as being anticipated by *Meyerzon*. The Applicants disagree and traverse.

As background information, the Applicant provides the following summary of salient aspects of *Meyerzon*.

Meyerzon uses two distinct methods for processing a newly crawled document that is determined to be a duplicate of a previously crawled document. In one method, *Meyerzon* ignores the new document, because the document database already has the document. In the second method, if the new document is determined to have updated content, the new document is processed and replaces the old document. *Meyerzon* has no need to track "duplicate documents", and in fact describes no data structures for keeping track of "duplicate documents".

Independent claim 12 requires use of document ranks to update information about a set of duplicate documents:

> ... each table ... storing information identifying
> documents having a same document identifier ...
>
> updating the information stored in at least one of the
> tables in accordance with the document ranks of the identified
> set of documents and the newly crawled document; ....

The term "rank" is used in *Meyerzon* only once (column 2 lines 3-16) in the background section to describe operation of a search engine, as opposed to a web crawler. As it happens, the "rank" in *Meyerzon* is a query dependent ordering of search results, as opposed to the query independent score described in the present application (e.g., paragraph 0007 on page 2). More importantly, *Meyerzon* does not use document rank, regardless of definition, for use in detecting or processing duplicate documents, and therefore *Meyerzon* does not teach or anticipate claim 12.

Claim 12 also describes

> reading information stored in the plurality of tables to
> identify a set of documents, if any, sharing the document
> identifier of the newly created document; ....

In other words, documents thought to be duplicates are stored in a series of tables. *Meyerzon* (column 9, lines 18-29) does not store any information about documents thought to be

duplicates. In one method, when the *Meyerzon* crawler encounters a document that already exists in the History Table, it ignores the document (column 9, lines 18-29). Similarly, in the incremental update method of *Meyerzon, Meyerzon* does not read information to identify a set of documents that share the same document identifier. Instead, *Meyerzon* is only interested in a yes/no answer to the question of whether it has previously crawled any document having the same document identifier. Restated, in *Meyerzon,* there is no "identified set of documents." Because *Meyerzon* keeps no record of any newly acquired documents thought to be duplicates, and because *Meyerzon,* does not read information in tables to identify a set of duplicate documents, it does not teach or anticipate claim 12.

Claim 12 further describes "determining a <u>representative document</u> for the newly crawled documents and <u>the identified set of documents</u>". *Meyerzon* (column 9, lines 32-40) does not identify a representative document, and in fact, ignores the information in the newly crawled document when it is determined to be a duplicate. Furthermore, as noted above, in *Meyerzon,* there is no "identified set of documents" as required by the above-quoted portion of claim 12. Because there is no comparison between the old duplicate document and newly crawled document, no representative document is selected and, therefore, *Meyerzon* does not teach or anticipate claim 12.

For these reasons The Applicants respectfully requests that the Examiner withdrawal rejection of claim 12.

Dependent claim 14 describes

> "comparing the document rank of <u>the newly crawled</u>
> document with that of <u>a document in</u> the identified set in
> accordance with a set of predefined comparison criteria;
> selecting the <u>newly crawled</u> document as the representative
> document if the set of predefined comparison criteria are met".

Meyerzon (column 5, lines 20-40) describes a method for identifying documents that have changed since a previous search using a crawl number. The identification of recently modified documents is entirely different than the comparison of two duplicate documents described in claim 14. For this reason *Meyerzon* does not teach or anticipate claim 14.

Independent claims 40 (as amended) and 50 are all patentable over *Meyerzon* for at least the same reasons as claim 12. Furthermore, the claims depending from claims 12, 40 and 50 are also patentable for at least the same reasons as those described above.

Claim Rejections - 35 U.S.C. § 103

The Examiner has rejected claims 18-20, 37-39 and 56-58 under 35 U.S.C. 103(a) as unpatentable over *Meyerzon* in view of *Rujan*. To establish a prima facie case of obviousness:

> The prior art reference (or references when combined) must <u>teach or suggest</u> <u>all the claim</u> <u>limitations</u>. The teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art and not based on applicant's disclosure.[1]

Independent claims 18, 37 and 56 specify:

> ...a plurality of tables ... storing information identifying documents having a same document identifier...
> ...and each identified document having an associated document rank...
> ...updating the information stored in the current table in accordance with the document rankings of the identified set of documents and the newly crawled document;
> determining a representative document for the newly crawled document and the identified set of documents...

As described above, *Meyerzon* does not assign a query independent rank to duplicate documents. When a duplicate document is detected, *Meyerzon* provides no means for updating information in the history table <u>based on document rank</u>. If the newly crawled document is determined to be a duplicate, *Meyerzon* simply ignores the content of the newly crawled document and leaves the history table and index table entry intact. Because there can only be one duplicate document (i.e., only one document having a particular document identifier) in the history file, *Meyerzon* can not collect multiple duplicate documents in a plurality of tables. Thus, *Meyerzon* does not "store information identifying documents having a same document identifier." Furthermore, the method described in *Myerson* does not allow for the selection of a <u>representative document</u> from among "the newly crawled document and the identified set of documents," at least in part because in *Meyerzon* there is no "identified set of documents".

For these reasons, *Meyerzon* in view of *Rujan* does not teach all the limitations described in claims 18, 37 and 56. Furthermore, the claims depending from claims 18, 37, and 56 are also patentable for at least the same reasons as those described above. In light of

---

[1] *In re Vaeck*, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991).

these arguments, the Applicants respectfully request the Examiner withdrawal the rejection of claims 18-20, 37-39 and 56-58.

Claim Rejections - 35 U.S.C. § 103

The Examiner has rejected claims 49 under 35 U.S.C. 103(a) as unpatentable over *Meyerzon* in view of *Lambert*.

It is noted that Lambert does not teach storing information identifying a set of documents having the same document identifier, does not teach updating such information based on document rankings, and does not teach selecting a representative document from among a newly crawled document and an identified set of documents. Therefore independent claim 40 (as amended) and all its dependent claims (including claim 49) are patentable over the combined teachings of *Meyerzon* and *Lambert* for at least the same reasons as claim 12 described above.

CONCLUSION

In light of the amendments to the claims, the arguments presented above, and the terminal disclaimer, Applicants respectfully request that the Examiner reconsider this application with a view towards allowance. The Examiner is encouraged to call the undersigned attorney at (650) 843-4000 should any issues remain unresolved.

Respectfully submitted,

/ Gary S. Williams /

Date: June 15, 2006   31,066

Gary S. Williams
MORGAN, LEWIS & BOCKIUS LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, California 94306
(650) 843-4000